HW4_STAT5000

Xingyu Chen

September 23, 2021

Due through Canvas on Friday September 24, 2021 @ 11:59 PM. Instructions for "theoretical" questions: Answer all of the following questions. The "theoretical" problems should be neatly numbered, written out or typed, and solved. There are 18 lettered parts or stand-alone questions, so self-grade should total 36.

Total Score: 31/36

1 Theoretical Questions

1. Individuals A and B play a sequence of chess games until one player wins 10 games. A wins an individual game with probability p, and B wins a game with probability 1 - p (i.e., there are no draws). Let X denote the number of games played.

(a) What are the possible values of X? Q1: 2/2
Answer:
at least 10 games, at most 19 games.

(b) Obtain an expression for P (X = x). Q2: 2/2
Answer:
$P(X = x) =$

$$\binom{x-1}{9} * p^9 * (1-p)^{x-1-9} * p = \binom{x-1}{9} * p^{10} * (1-p)^{x-10}$$

(c) Let p = 0.5. Find P(X = 12). Q3: 2/2
Answer:
$P(X = 12) = \binom{12-1}{9} * 0.5^{10} * (1-0.5)^{12-10} = 55 * 0.5^{10} * 0.5^2 = 55/4096 = 1.34\%$

2.

(a) A traffic office wishes to monitor the number of vehicles crossing a certain bridge in the city in any given day. What family of distributions will they most likely be observing? Q4: 2/2
Answer:
Poisson Distribution.

(b) If it is estimate that an average of 300 cars cross the bridge per day, what is the probability that less than 150 cars will cross the bridge on any given day? Q5: 1/2

Answer:

Poisson Distribution: Poi(x<150) = Poi(x<=149 | λ = 300) =

$$\sum_{i=1}^{150} \frac{300^i * e^{-300}}{i\,!}$$

= 0

3. A couple wishes to have exactly two female children in their family. They will have children until this condition is fulfilled. Assume that male and female births are equally likely.

(a) What is the probability that the family has x male children? Q6: 2/2

Answer:

There are x male children if there is one female among the first x+1 children, and a female on the (x+1)-th birth.

NB(r, p) = NB(x, 2) = $\binom{x+1}{1} * 0.5^2 * 0.5^x = \binom{x+1}{1} * 0.5^{x+2}$

(b) What is the probability that the family has four children? Q7: 2/2

Answer:

Pr(x=2) =

$$3 * (1/2)^4$$

= 3/16 = 0.1875

(c) What is the probability that the family has at most four children? Q8: 2/2

Answer:

Pr(x=2) + Pr(x=1) + Pr(x=0) =

$$0.1875 + 2 * (1/2)^3 + 1 * 0.5^2$$

= 11/16 = 0.6875

4. A certain lab machine has 22 rings. With each use, these rings can fail, and an oil leak occurs. The probability of any ring failing during machine use is 8%. The rings are independent, and the failure of one ring does not impact the probability of failing for other rings. The machine gets serviced after each use, so any damaged rings are repaired after each use.

(a) If 4 or more rings fail, the entire machine will shut down. What is the probability of the machine shutting down on any one use? Q9: 2/2

Answer:

X ~ Binomial(n = 22, p = 0.08)

P(shut down) = P (X >= 4)

$= \sum_{i=4}^{22} \binom{22}{x} 0.08^x 0.92^{22-x}$

= 0.094

(b) What is the probability that the machine runs successfully at least 5 times before shutting down? Q10: 2/2

Answer:

p(no shut down) = 1- p(shut down) = 1-$\sum_{i=4}^{22} \binom{22}{x} 0.08^x 0.92^{22-x}$ = 0.906

p(no shut down in first 5 uses) = p(no shut down)^5 = 0.610

5. Let X = the leading digit of a randomly selected number from a large accounting ledger. So, for example, if we randomly draw the number \$20, 695, then X = 2. People who make up numbers to commit accounting fraud tend to give X a (discrete) uniform distribution, i.e., P (X = x) = 1/9, for x ∈ {1, . . . , 9}. However, there is empirical evidence that suggests that "naturally occurring" numbers (e.g., numbers in a non-fraudulent accounting ledger) have leading digits that do not follow a uniform distribution. Instead, they follow a distribution defined by:

$$f(x) = \log_{10}(\frac{x+1}{x}), x = 1, 2, ..., 9.$$

(a) Show that f (x) = P (X = x) is, in fact, a probability distribution function. Q11: 2/2

Answer:

The first digits of the numbers are uniformly and randomly distributed. The probability of digits is proportional to the space between x and x+1 on a logarithmic scale.

(b) Compute the individual probabilities for x ∈ {1, . . . , 9}, and compare them to the corresponding discrete uniform distribution (i.e., P(X = x) = 1/9). What do you notice? Q12: 1/2

Answer:

f(1): 30.1%
f(2): 17.6%
f(3): 12.5%
f(4): 9.7%
f(5): 7.9%
f(6): 6.7%
f(7): 5.8%
f(8): 5.1%
f(9): 4.6%

The corresponding discrete uniform distribution (i.e., P(X = x) = 1/9) is evenly distribute. However the individual probabilities for x is based on log x so it is uniformly and randomly distributed with 1 is the most common leading digits.

(c) Obtain the cdf of X. Q13: 2/2

Answer:

$F_x(x) = Pr(X <= x)$
$= \sum_1^x \log_{10}(1 + 1/i)$
$= \sum_1^x (\log_{10}(i + 1) - log_{10}(i))$
$= \log_{10}(x + 1) - \log_{10}(1)$

$= \log_{10}(x+1)$

(d) Using the cdf, what is the probability that the leading digit is at most 4? At least 5?
Q14: 2/2
Answer:
P(x <= 4) = 0.602

P(x >= 5) = 1 - p(x <= 4) = 0.398

2 Computational Questions

Instructions for "computational" questions: Your work should be neatly done and in- clude all graphs, code, and comments, labeled and in order based on the problem you are addressing. Do not put graphs in at the end, stick code in random locations, or do anything else that will make this homework difficult to read and grade.

1. Let X follow a geometric distribution, where the probability of success is 0.2.

(a) What is the probability that 7 failures occur before the first success? Q15: 2/2
Answer:

```
dgeom(7, 0.2, log = FALSE)
```

```
## [1] 0.04194304
```

(b) What is the probability that 7 or more failures occur before the first success? Q16: 1/2
Answer:

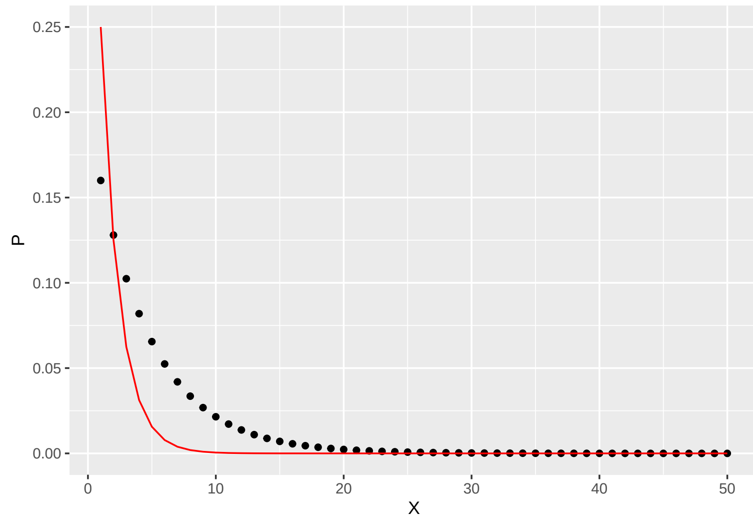
```
sum = dgeom(6, 0.2, log = FALSE)
sum = sum + dgeom(5, 0.2, log = FALSE)
sum = sum + dgeom(4, 0.2, log = FALSE)
sum = sum + dgeom(3, 0.2, log = FALSE)
sum = sum + dgeom(2, 0.2, log = FALSE)
sum = sum + dgeom(1, 0.2, log = FALSE)
sum = sum + dgeom(0, 0.2, log = FALSE)
1 - sum
```

```
## [1] 0.2097152
```

(c) Plot this pdf for X = 1, . . . , 50, using the X values on the x-axis, and P (X = x) on the y-axis. To this same plot, add a red line showing the probabilities for X = 1, . . . , 50 when the probability of success is 0.5. What do you notice? Q17: 1/2
Answer:

```
library(tidyverse)
x_dgeom <- seq(1, 50, by = 1)
y_dgeom <- dgeom(x_dgeom, prob = 0.2, log = FALSE)
y_dgeom_sub <- dgeom(x_dgeom, prob = 0.5, log = FALSE)
```

```r
pdf_1 <- data.frame("X" = x_dgeom, "P" = y_dgeom)
ggplot(pdf_1, aes(x = X, P)) + geom_point() +
geom_line(aes(y = y_dgeom_sub), color = "red")
```



The greater the success of probability, the greater probability of P(X = x) when x is small but the probability drop more quickly when x become greater.

2. Run a simulation that estimates the probability calculation in theoretical question 1 (c). Q18: 1/2

   Answer:

```r
dbinom(9,11,0.5,log=FALSE) * 0.5
```

```
## [1] 0.01342773
```